AFOSR 69-0405TR

AD 683 113

F44620-69-C-0068

1. This document has been approved for public release and sale; its distribution is unlimited.

THE SHUFORD-MASSENGILL CORPORATION

SM

31

CONFIDENCE TESTING AT THE OFFICER TRAINING SCHOOL, LACKLAND AIR FORCE BASE: SEPTEMBER 1968

Emir H. Shuford, Jr. and H. Edward Massengill

Sponsored by

Advanced Research Projects Agency
ARPA Order No. 833

CONFIDENCE TESTING AT THE OFFICER TRAINING SCHOOL, LACKLAND AIR FORCE BASE: SEPTEMBER 1968

Emir H. Shuford, Jr. and E. Edward Massengill

BACKGROUND

On the morning of 13 September 1968, the Air Training Command of the United States Air Force conducted a preliminary tryout of *Valid Confidence* testing by using the method to administer an achievement test to 98 officer candidates of the Officer Training School (OTS), Lackland Air Force Base, Texas. The experiment was observed by Major Donald W. Jones and Captain J. R. Schwille of Headquarters, Air Training Command, while Dr. Shuford and Mr. Massengill of The Shuford-Massengill Corporation assisted Mr. Anthony P. Barra, Director of Testing at OTS, in the administration of a unit test concerned with the leadership curriculum at OTS. This test, with alternate forms designated L1-2 and L1-2A, was a revised version of the test previously taken by these same officer candidates as part of the normal course of instruction and evaluation at OTS.

The original intent had been to readminister a test that had been taken the day before by these officer candidates. It developed that the officer candidates had been briefed on the results of the test and by that morning had almost complete confidence in the correct answers to all the questions. Since administering a test under these conditions would have not required students to demonstrate that they could discriminate according to the quality of information available to them, it was decided to readminister the leadership test which had been given some weeks earlier. By this time, of course, it was to be expected that the students had forgotten some of the correct answers and, in some instances, would not be justified in having complete confidence in an answer.

PROCEDURE

A tape recording was used to instruct the students on how to take a *Valid Confidence* test. After playing the tape, Mr. Massengill answered some questions from the students and the students took a short practice test using the *SCoRule* and one of the answer sheets. Instructions and a short break took approximately one hour. The students then returned to the auditorium and responded to the 58 four-alternative multiple-choice test items using the *SCoRule* and three additional answer sheets. The students were allowed one and one-half hours to complete the test. This is the time normally allowed at OTS for the administration of a unit test.

On completing the test, the student wrote the time on his answer sheet and passed in the test booklet, his answer sheets, and the *SCoRule*. The distribution of finishing times is shown in Figure 1. It is apparent that about 4/5 of the students finished well ahead of the time limit and about 1/5 required the complete period to complete the test.

The distribution of finishing times for the students taking the test as a choice test is not available to us, but even without this comparative data some conclusions can be reached. Even though it must be true that it does take longer

to write down a degree of confidence for each of the answers than to choose among the answers and to indicate this choice, and even though confidence testing tends to make students think more carefully about test questions, the data shown in Figure 1 indicates it is quite feasible to give a *Valid Confidence* test within the time limits usually allowed for choice tests. This is so because most of the students finished early and had time left over before the time limit was reached. In shifting from choice to confidence testing, these students would take somewhat longer to complete the test and as a consequence would have less time left over before exceeding the time limit while the 1/5 of the class which used the full time to finish the test might act the same no matter what method of testing is used, or amount of time provided. Notice that the distribution of finishing times clusters around 50 to 60 minutes and tails off to shorter and longer times. The students taking the full time to finish the test seem not to be a part of this distribution, that is, these students do not represent a truncation at the tail of the distribution as evidenced by the fact that there is a gap of 15 minutes between the tail of the main distribution and the time limits of the test.

Thus, if these students really understood the instructions and realized what confidence testing is all about, then we can say that, with a one time investment of one hour instructing students on how to take a test this way, it is possible to administer a test as a *Valid Confidence* test allowing no more time for this administration than was previously allowed for administration of the test as a choice test.

DID THE STUDENTS UNDERSTAND THE INSTRUCTIONS?

The test data was analyzed by The Shuford-Massengill Corporation using test keys provided by the Officer Training School, Lackland Air Force Base. As described in Shuford & Massengill (1968) and Shuford (1969) there is a basic test for the meaning and validity of confidence which can serve to indicate whether or not the students understood the test instructions for the *Valid Confidence* testing procedures. The short form of this validity test is to see if the percent "Z" answers correct characteristic of the student is indeed greater than the expected percent correct answers inferred for his taking the test as a choice test. These two statistics can be computed for each student by counting the total number of times the student placed complete confidence in an answer by assigning a "Z" to it and then finding the percent of times that these answers were actually correct answers in order to obtain the percent "Z" answers correct. The inferred expected correct answers may be found by using his confidence responses to infer what choice the student would have made if the test had been administered as a choice test. (The underlying assumption here is that the student would have chosen that answer in which he had the greatest amount of confidence.) This is easily determined except for those instances in which two or more answers are tied for highest confidence. In these cases, the student is given an expected item choice score. For example, when the correct and one other answer is tied for the highest confidence, then the student is given 1/2 of a point while if the correct answer and two other answers are tied for highest confidence, the student is given 1/3 of a point, and so on. These inferred item choice scores of either one full point, 1/2 point, 1/3 point, 1/4 point, or zero points are then summed to obtain a raw choice score and then divided by the total number of items (in this case, 58) to obtain the inferred expected correct answers as plotted in Figure 2.

Notice that the data for each of the 98 students passes this test, as indicated by all points falling above the diagonal line, i.e., the confidence responses yield more information than do the choices. Notice also that there are truly great individual differences in this measure ranging from the one student who

made the top choice test score but whose confidence test data yielded very little more information than did his choice test data up to the 13 students who evidenced perfection in their use of the "Z" response in *Valid Confidence* testing. In summary, Figure 2 indicates that the instructions were learned well by at least a great majority of the students.

EXTERNAL VALIDITY ANALYSIS

To further understand the implications of this validity test, it is necessary to examine in detail the test data for some of these students, particularly those at the extremes. The five students whose data points are circled in Figure 2 and who fall nearer the diagonal line are students whose confidence data is not telling us much more than would their choice test data, while the five students whose choice data points are circled and appear at the 100% level up near the top level of Figure 2 are students whose confidence test data are telling us much more than would their choice test data. This gives two extreme groups of students.

We can do a full analysis of the test data of these students by finding a percent correct for each possible assignment of degree of confidence and plotting this as shown in Figures 3 and 4. For each graph the data has been averaged for the group of five students. The dashed line has been derived from the inferred choice test data and indicates what would happen if these students had given no more information in their confidence responses than in their choice responses while the diagonal line represents perfection.

Figure 3 shows the average behavior for the five students whose confidence responses yielded minimal gain in information over choice testing. The empirical function (represented by the data points connected by the bold straight line) does indeed have a steeper slope than does the dashed line thus, indicating that even the students at this extreme are giving more information with their confidence responses. Notice that the data points fall fairly close to the diagonal line except for extremely high and extremely low degrees of confidence. Therefore, over the middle of the range of confidence these students fairly realistically evaluate the quality of the information available to them. They neither overvalue nor undervalue the confidence jsutified by the information at hand. When, however, the situation is such that a fairly high degree of confidence is justified, these students tend to "go all the way" and place 100% confidence on the answer rather than the 80% or 90% which is probably justified. Likewise, at the other extreme, when the information is such that a student can almost exclude the answer as a logical possibility the students again "go all the way" and put 0% confidence on the answer rather than the 5% or 10% which probably is justified.

Now look at Figure 4 which shows the average data for five of the students whose confidence responses yield maximal gain in information over choice testing. The empirical line varies both above and below the diagonal line due undoubtedly to the random fluctuation resulting from the small sample sizes yielding the data points. A theoretical function fitted to these data points would undoubtedly be a straight line with a slope very close to one and almost identical to the diagonal line representing perfection in evaluating the quality of information. The data of these five students represents exceptional realism and deviates trivially from ideal performance. One wonders if the exceptional ability of these students in evaluating information also manifests itself in content areas other than leadership and in forms of behavior other than test-taking.

REALISM IN EVALUATING THE QUALITY OF INFORMATION

A student's skill at evaluating the quality of information as reflected in this type of analysis is a totally new ability measure available only from confidence testing. To the extent that this skill proves to be stable and characteristic of the individual over several domains of behavior its existence as an ability would be confirmed. This in turn might have far reaching implications. For example, this ability is different from how much information or knowledge an individual possesses. He can possess a great deal of information but be very poor at evaluating its quality. Or he can possess very little information but be quite expert at evaluating the quality of this information and, of course, vice versa. No matter how much information an individual possesses, however, his ability to use this information effectively to make decisions of high quality remains limited by his ability to evaluate the quality of the information.

To see this most clearly, suppose that we were having officers evaluate the quality of certain intelligence information and that this evaluation was done in terms of degree of confidence as to the existence of certain situations. These degrees of confidence are then fed as probabilities into an information system (possibly computer based) which combines these probabilities with the utilities of the possible outcomes according to the rules of mathematical decision theory in order to recommend a course of action. This decision system applies logic and mathematics to the data at hand to make the best possible decisions. The effectiveness of these decisions, therefore, would only be limited by the data that the system receives in terms of the degrees of confidence. The value of this data depends not only on the information available to the officer but on the realism with which he evaluates this information. This latter factor is exactly the measure that we are dealing with in *Valid Confidence* testing.

Consider two officers having exactly the same information making inputs to such a system. Suppose one officer is nearly perfect in evaluating the quality of information while the other officer cannot tell good from bad information. He doesn't know what he knows and he doesn't know what he doesn't know. Clearly the decision system would perform much more effectively with inputs from the first officer who was able to give realistic values to the information.

We don't need to have a computer-based decision system to make this argument valid. The officer could very well be making his own decisions and the complete system would be internal to the officer. It doesn't really matter. The officer can still behave in accord with the logic and mathematics of decision theory and the same limitations would apply. The effectiveness of his decisions would be limited by his ability to evaluate the quality of information. An officer's decision-making performance could be improved by giving him more information or by improving his ability to evaluate information. In many instances decisions have to be made in situations where there is no possibility of getting additional information. In these situations, the best we can do for the officer is to make sure that he is able to realistically evaluate the information to get the most out of the information at hand. Thus, teaching officers to realistically evaluate the quality of information may become a major educational and behavioral objective and might benefit the students' performance not only at Officer Training School but throughout his career both in educational and operational settings.

The data from this one test administration clearly indicate that there are wide individual differences among the officer candidates in their ability to evaluate the quality of information. The possibility exists that these differences may be quite temporary because of different understandings of the test instructions.

Experience, however, in a public school setting where students are tested weekly or more often indicates that these individual differences do remain stable over a long period of time. Experience also indicates that certain techniques can be successfully employed to improve the realism with which many of the students evaluate information. It would seem worthwhile to investigate the possibility that this is an ability characteristic of every individual and further, that it is an ability which can be taught and improved upon with practice.

COMPARISON OF TOTAL TEST SCORES

Remember that these students had taken a previous version of this test some weeks prior to the experimental administration of L1-2 and L1-2A. It was taken as a choice test as part of the normal instructional and evaluation program at OTS. The records were retrieved for each student so that his score could be compared with his score from this experimental administration. Figure 5 shows this original test score plotted against inferred choice score for each of the 98 students in the experimental group. Examination of Figure 5 shows that the scores are indeed correlated but not too highly. As would be expected, the average test score was much higher for the original administration of the test than the readministration some weeks later.

The original test score is compared with the *Valid Confidence* score in Figure 6. As before, there is a correlation but not too high a one between these two sets of test scores. The positive correlation in both of these Figures indicates that the original test and the experimental test are measuring some things in common. This would certainly be expected and would be a minimal requirement for any new testing method.

A more revealing comparison is to look at the association between the inferred choice score and the *Valid Confidence* score for the experimental administration. This relation is not obscured by retention, selection of test items, etc. Examination of Figure 7 indicates that the inferred choice score and *Valid Confidence* score are indeed related with a correlation higher than before but the association is far from perfect. The *Valid Confidence* score for every officer candidate is higher than his inferred choice score. Accepting Massengill's (1969) argument that the *Valid Confidence* or information score is the measure of the amount of information demonstrated by the student with respect to the test and, thus, the fair way to assess the student, it becomes apparent that choice testing underestimates the amount of information that the students demonstrate. In addition, the choice test makes many errors in ranking the students according to their demonstrated knowledge. For example, the student making the highest *Valid Confidence* score (thus indicating the possession of more information than anyone else in the class) is tied with four other people for a rank of 5.5 according to the choice score. For another example, the student making a choice score of 41 who is 14th in rank according to *Valid Confidence* score finds that if he were taking the test as a choice test, 44 students who demonstrated less information than he did would have been given higher choice scores and that the choice test would have ranked him as tied with 12 other students who in fact demonstrated less information than he did. In brief, assessing the accomplishment of students according to the *Valid Confidence* score can make quite a difference.

TEST SCORE AND FUTURE PERFORMANCE

There are many ways to understand why the *Valid Confidence* score is more valid in assessing students than is the choice score. Massengill (1969) relates it to the measure of quantity of information and shows many concrete examples of how the information score serves to eliminate the operation of chance in guessing and how it rewards the student who is uninformed more than the student who is misinformed. Another approach is to consider the nature of the test itself and the relation of the knowledge demonstrated on the test to the performance of the student in situations outside of the test administration. If the test is made up of independent bits and pieces of knowledge and the performance of related tasks outside of the test situation is such that performance depends upon how many of these bits and pieces of knowledge are mastered, then a choice type test score is appropriate. To be more explicit, if a "real world" task can be performed just as well if the student has bits of knowledge represented by test items 1, 2, and 3 as if he had bits of knowledge represented by test items 1, 2, and 4 then we say that the knowledge tested by the items is substitutable and the performance of the person depends upon how many bits of knowledge he has acquired. The more bits of knowledge he has acquired, the better he is able to perform the task. This is the type of situation that is best assessed by a choice test score.

We need to distinguish however, another type of relation between the knowledges assessed by a test and performance in a "real world" situation. Many "real world" situations seem to have the characteristic that in order to perform them at all successfully you need to have certain items of information and if you have not mastered all these items, you cannot perform the task. In particular, if you are misinformed on one item of information, then you are guaranteed to do the task wrong. The items of information cannot be substituted one for another. Now certainly some tasks are more complex than others and they require the mastery of more items of information than do the other tasks. For an approximation we can say that being misinformed on one item of information so damages the performance of the individual that he must have completely mastered several other items of information in order to make up for it in his general behavior. This assumption is implied in the *Valid Confidence* score and is reflected in the fact that *Valid Confidence* testing much more severely penalizes the student who has complete confidence in a wrong answer and denies the logical possibility of the correct answer than it rewards the student who has complete confidence in the right answer. The score received by the uninformed student is much closer to full credit than to no credit.

In summary, if the information structure assessed by a test has the characteristic that different parts of it must be put together for the successful performance of a task in a "real world" setting, the *Valid Confidence* score is clearly more appropriate than a choice score. If the information structure assessed by a test is composed of unrelated bits and pieces of knowledge then a modification of the *Valid Confidence* score might be more appropriate. In the particular case of L1-2 and L1-2A, and in fact in several of the other tests administered at the Officer Training School, these are really test batteries composed of subtests, each subtest measuring a particular teaching objective. A criterion score is placed on each teaching objective so that the student must make a passing score on each subtest. If not, the student must study again for the test and take that teaching objective over again. If a number of teaching objectives are failed, the student fails the test and possibly the whole course.

COMPARISON OF SUBTEST SCORES

This policy makes it important to compare *Valid Confidence* scores and inferred choice scores for the teaching objectives on L1-2 and L1-2A. This has been done

for the first two teaching objectives. Figure 8 shows the association between the choice and confidence scores for the seven-item subtest measuring mastery of Teaching Objective No. 1, while Figure 9 shows the same for Teaching Objective No. 2. These data yield ample evidence that *Valid Confidence* scores are not the same as choice scores and that different students would be passed or failed under the two grading systems.

For the highest possible choice score of seven correct out of the seven items, the *Valid Confidence* score is of course less than or equal to the choice score of seven. Most of the students in this category make a somewhat smaller *Valid Confidence* score than seven. This is so because these students indicate that they have less than complete confidence in the correct answer on one or more of the seven test items. To the extent that they have less than complete confidence in the correct answers this score must fall below that of a student who has complete confidence in all the answers. At the other extreme of the students making low choice scores of two and three correct out of the seven items, all made considerably higher *Valid Confidence* scores. This tends to be so because these students realized that they did not know the answer to some of the questions and so indicated whereas if they had indicated complete confidence in an incorrect answer, their choice score and *Valid Confidence* score would have been the same. *Valid Confidence* testing is rewarding them for knowing that they don't know.

In the normal use of this test, a choice score of five or more correct out of the seven items meets the teaching objective. Although normally the students would score better, it is still interesting to look at how many and which students would fail according to the choice score and which students would fail according to the *Valid Confidence* score. In Teaching Objective 1, 23 students made a choice score of less than five and thus would fail. If we use the same criterion score of five for the *Valid Confidence* score, we find that 19 students would fail. Six of these 19 students, however, would pass the choice test, while the *Valid Confidence* score passes 10 students who fail the choice test.

In Teaching Objective 2, the choice test would fail 24 of the 98 students while only 17 of the students would fail the *Valid Confidence* test. Of these 17, two would pass the choice test but nine who pass the *Valid Confidence* test would fail the choice test.

In summary, on the teaching objective subtests, students tend to make higher *Valid Confidence* scores than choice scores. Although choice score and confidence score are related, the relation is such that different students are passed and failed according to which method of scoring is used. These different instructional decisions are not due to changes from test to retest or other sources of variability inherent in the methods of scoring. From many points of view, the *Valid Confidence* score can be shown to be the fair way of assessing knowledge. If it is accepted that the information score is fair, the choice score cannot be. And to the extent that the choice score is passing and failing different students, choice testing is making serious errors in the assessment of students.

DIAGNOSIS OF STUDENT ACHIEVEMENT

Valid Confidence testing can be used to obtain a detailed diagnosis of each student's strengths and weaknesses with respect to the subject matter of the test. This was done by finding the student's state of knowledge based on his allocation of confidence among the possible answers and the results are shown in Table 1. The use of choice testing for this purpose would result in numerous errors as shown in Row IV of this table. Row V indicates the extent to which guessing was eliminated through the use of *Valid Confidence* testing.

SUMMARY AND CONCLUSIONS

A multiple-choice test on leadership was administered to 98 officer candidates in residence at the Officer Training School, Lackland Air Force Base. This test was administered using the materials and methods of *Valid Confidence* testing. Less than one hour was devoted to instructing the officer candidates on how to take a *Valid Confidence* test, and the normal time was then allowed for the students to respond to the 58 test items.

Analysis of the data indicates that taking a *Valid Confidence* test requires no more time than normally allotted to test administration. All the officer candidates understood the instructions and gave confidence responses which yielded more information than choice responses would have. Wide individual differences were observed in the officer candidates' ability to realistically evaluate the quality of information. Since an officer's ability to make effective decisions can be limited by the realism with which he is able to evaluate the quality of information, it appears worthwhile to conduct further research to measure this ability and to develop techniques to improve this ability in the officer candidate student population.

Test scores yielded by *Valid Confidence* testing are related to but are not the same as test scores obtained from choice testing. Several lines of reasoning lead to the conclusion that the *Valid Confidence* or information score provides a fairer basis for assessment than does the choice score. Therefore, to the extent that choice testing passes and fails different students than does *Valid Confidence* testing, use of choice testing as the means of assessing and grading students leads to unfair grades and incorrect instructional decisions. Further use of *Valid Confidence* testing as part of the normal program of instruction and evaluation at the Officer Training School should give students, instructors, and those responsible for evaluation further insights into the assessment properties of the information score.

REFERENCES

Massengill, H. Edward (1969) *Continuing classroom use of confidence testing.* SMC R-16.

Shuford, Emir H., Jr. (1969) *Systems of confidence weighting: theory and practice.* SMC R-15.

Shuford, Emir H., Jr. and Massengill, H. Edward (1968) *Airman Qualifying Examination-65 administered as a confidence test.* SMC R-12.

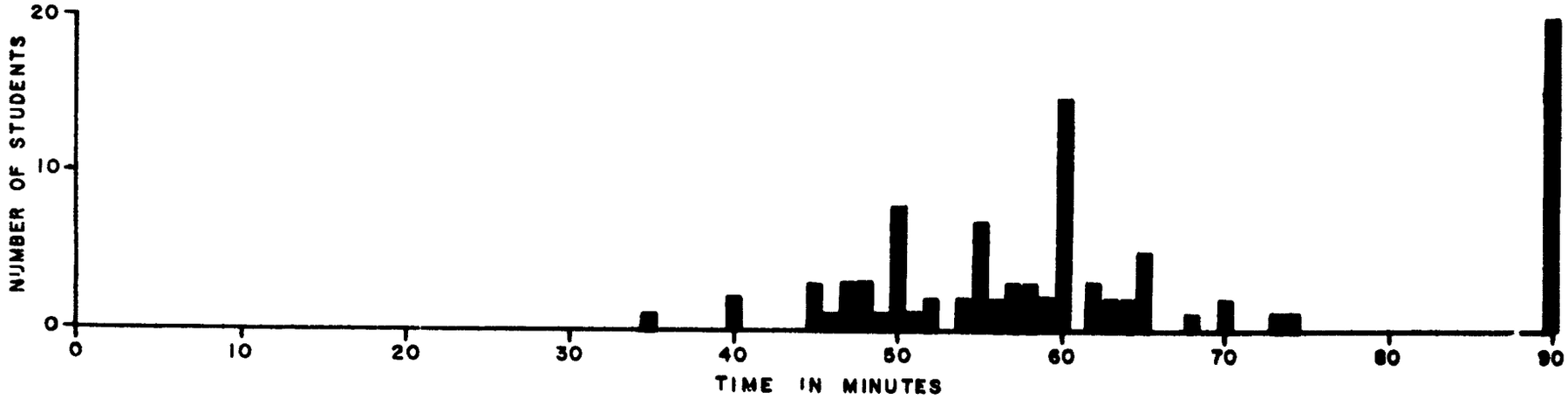

Figure 1. Distribution of time required to take LI-2 or LI-2A as a *Valid Confidence* test. Data based on 98 officer candidates. Ninety minute time limit specified at start of test.

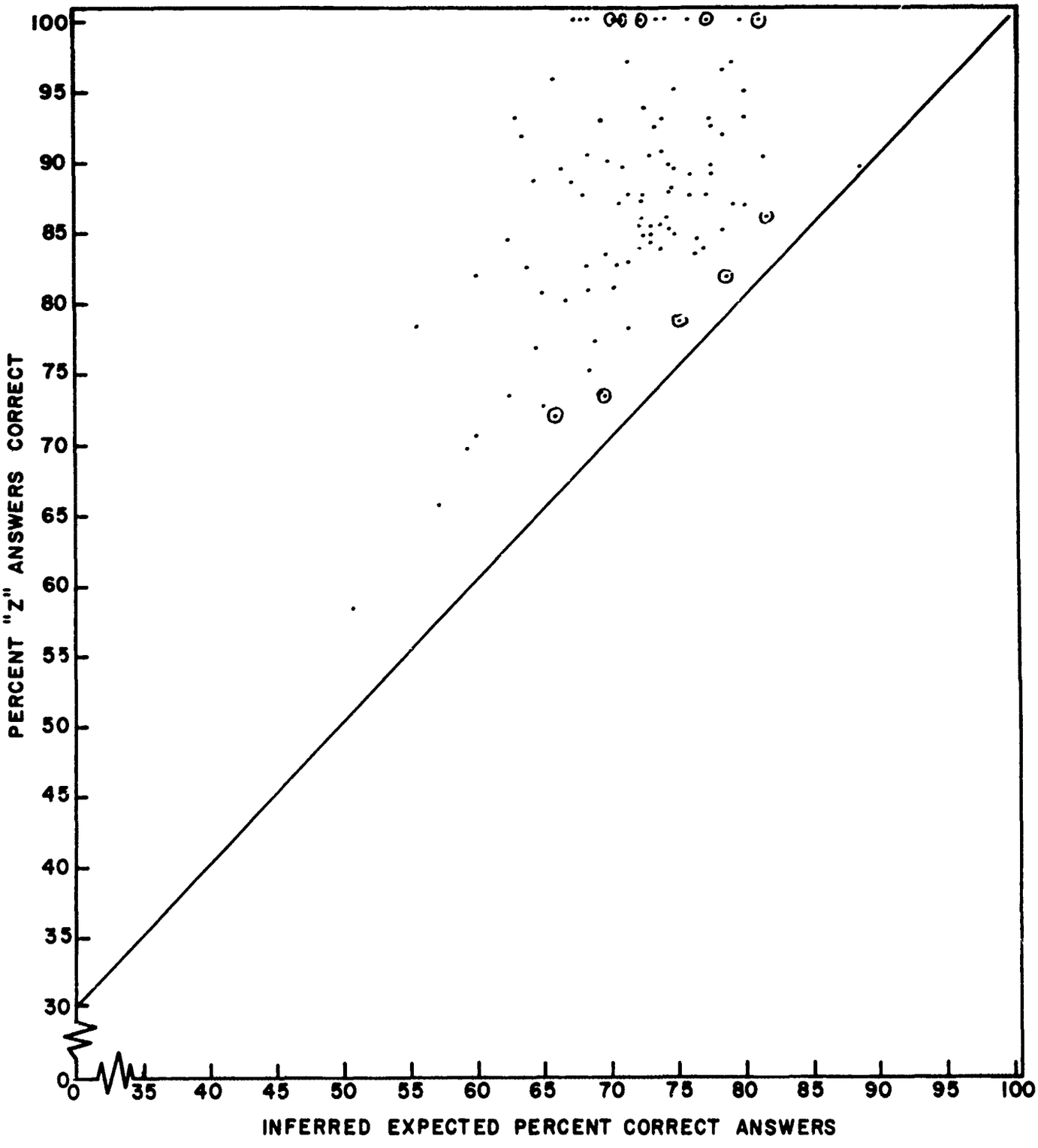

Figure 2. A measure of the fundamental validity and existence of confidence. A data point falling in the region to the left and above diagonal line indicates that student's confidence responses are yielding more information than would choice responses. Test data for the two groups of students whose data points are circled are analyzed further in Figures 3 and 4.

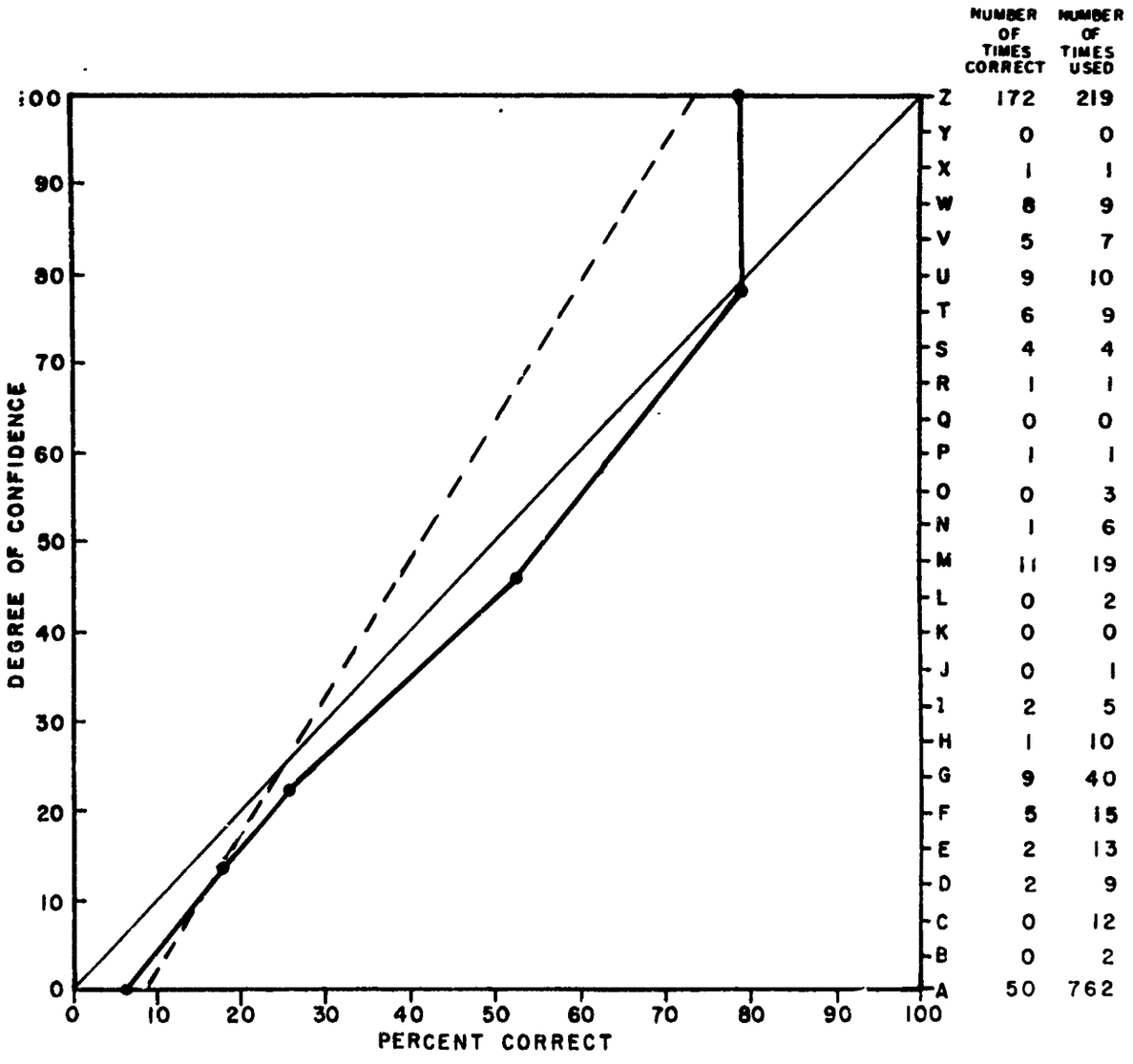

Figure 3. External validity graph based on the five students indicated in Figure 2, whose confidence responses yielded minimal gain in information over choice testing.

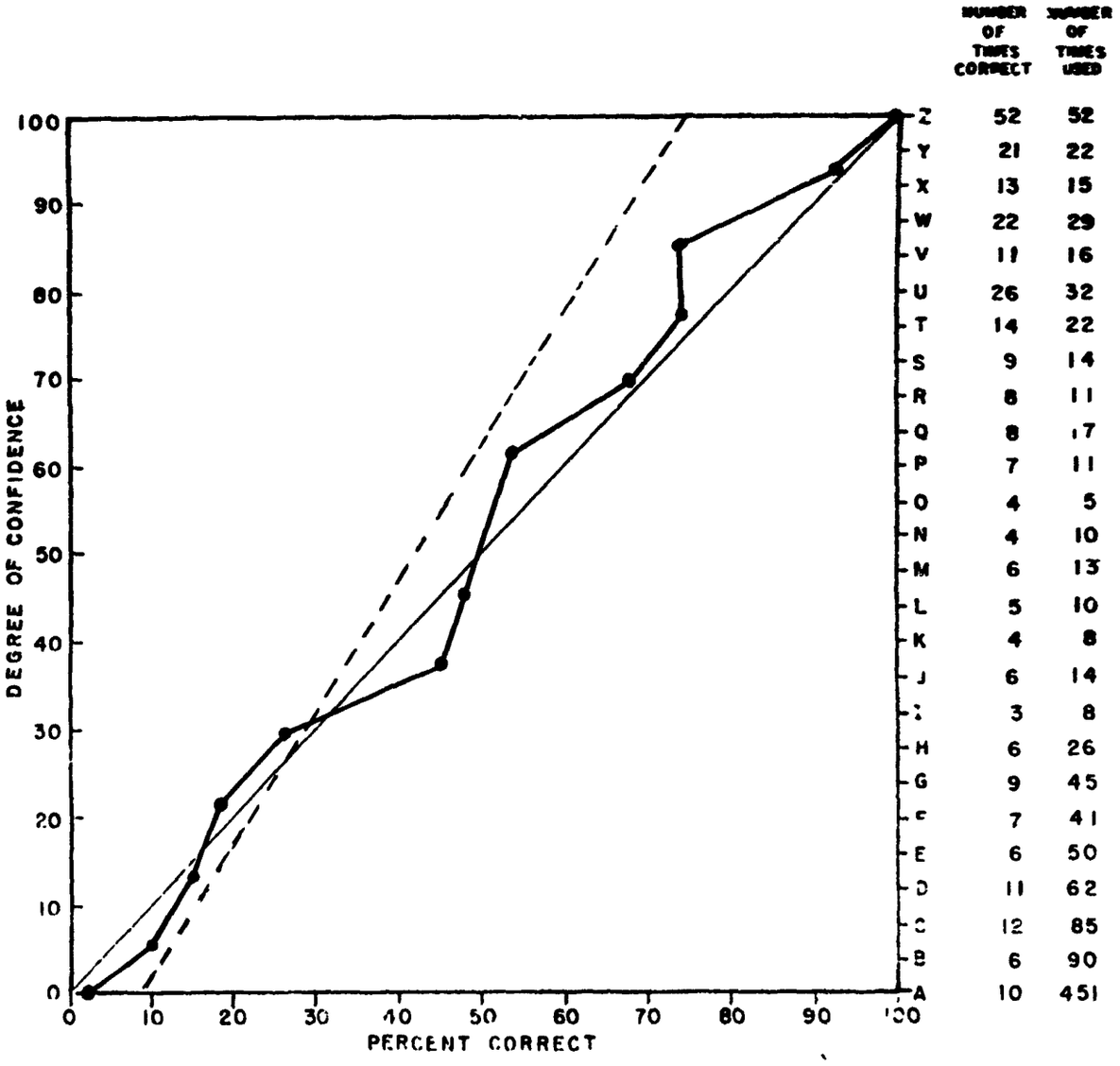

Figure 4. External validity graph based on the five students indicated in Figure 2, whose confiderce responses yielded maximal gain in information over choice testing.

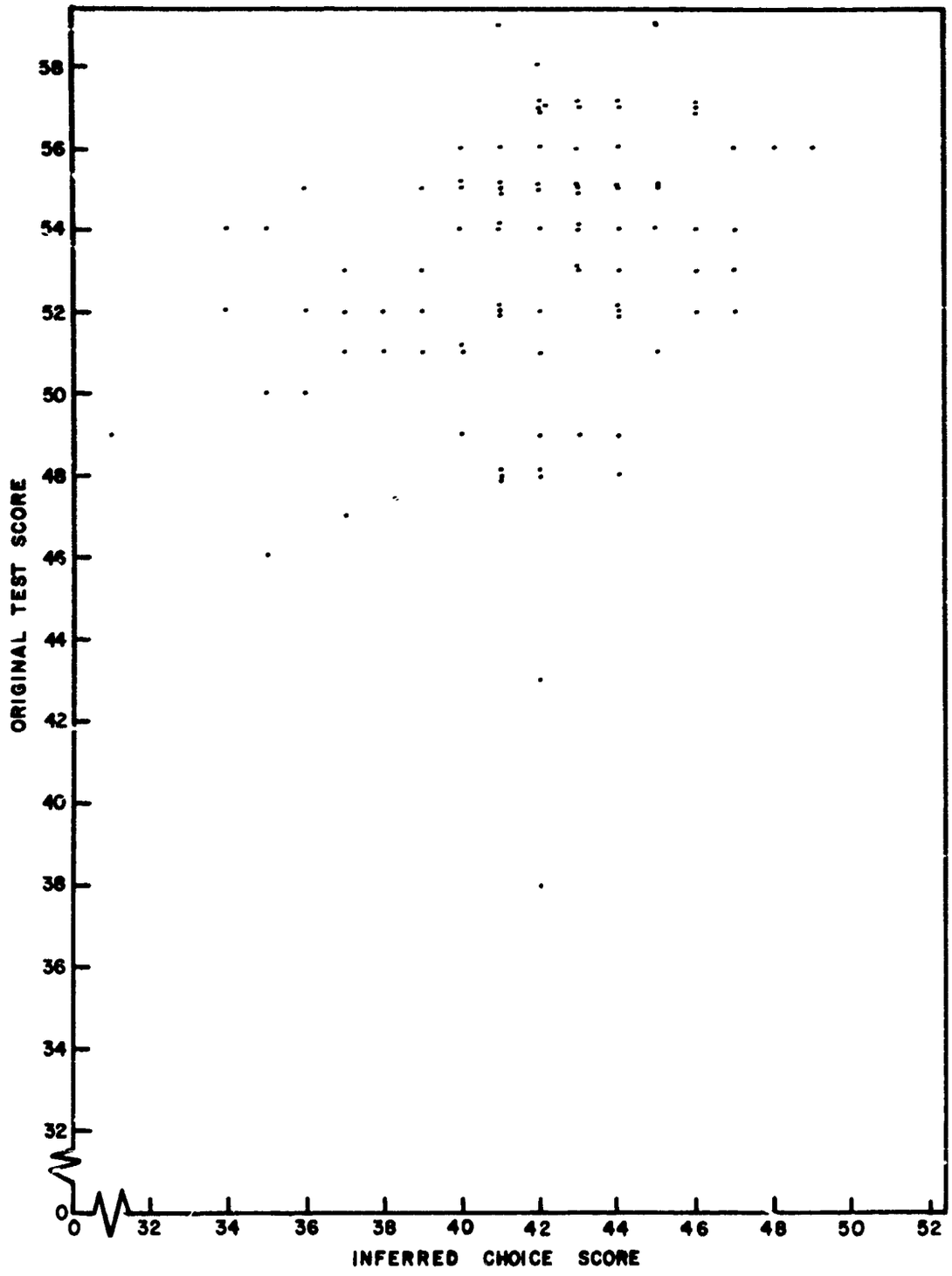

Figure 5. Association between Original Test Score, obtained from the earlier administration of a previous version of LI-2 or LI-2A as part of the routine testing program of the Officer Training School, and Inferred Choice Score obtained from the experimental administration of LI-2 or LI-2A as a *Valid Confidence* test.

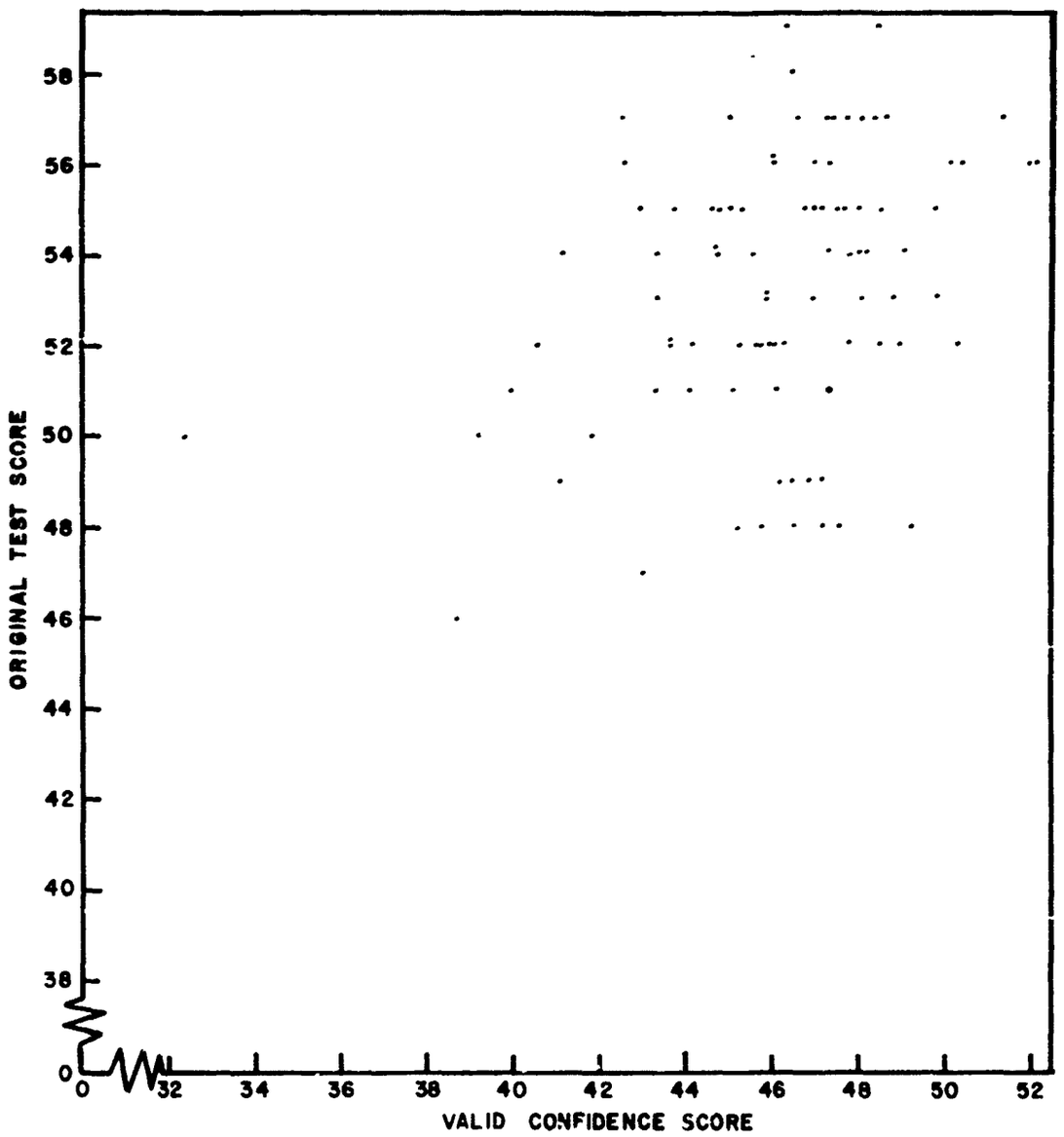

Figure 6. Association between Original Test Score, obtained from the earlier administration of a previous version of LI-2 or LI-2A as part of the routine testing program of the Officer Training School, and *Valid Confidence* Score obtained from the experimental administration of LI-2 or LI-2A.

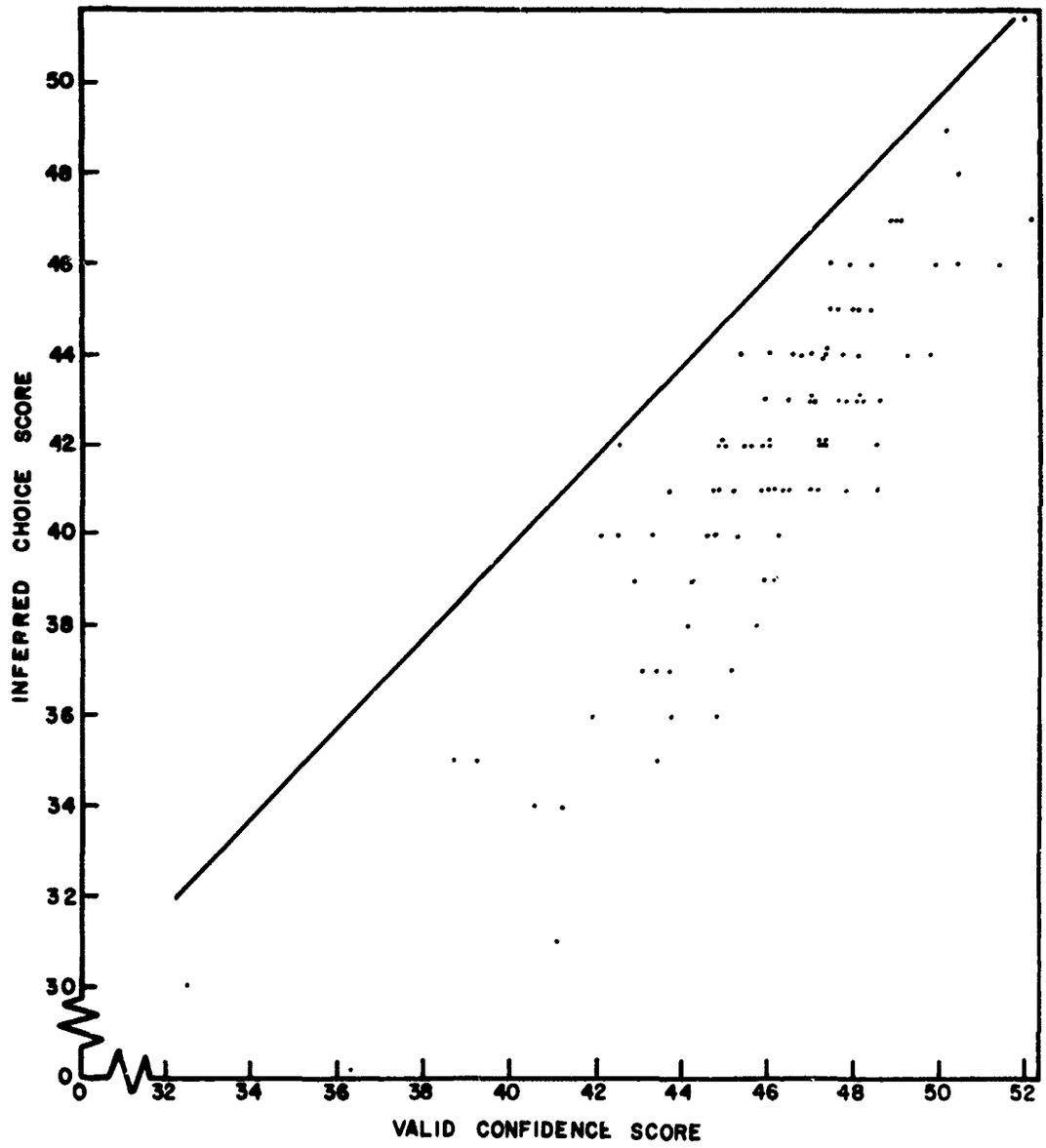

Figure 7. Association between the score the student would have made if LI-2 or LI-2A had been administered as a choice test, and the *Valid Confidence* score that the student actually received from the experimental administration.

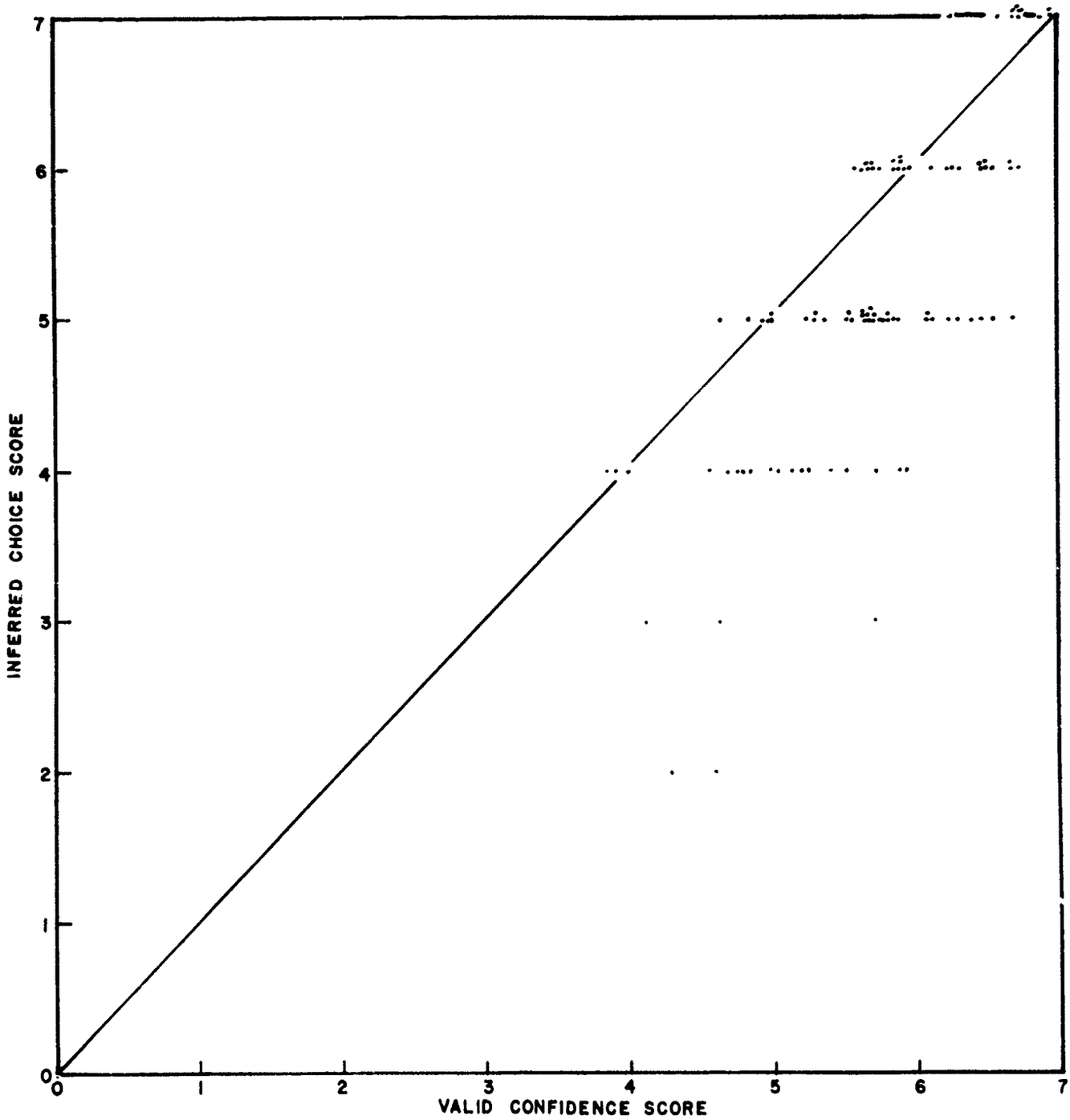

Figure 8. Association between choice and confidence scores for seven item subtest measuring mastery of Teaching Objective No. 1 on L1-2 or L1-2A.

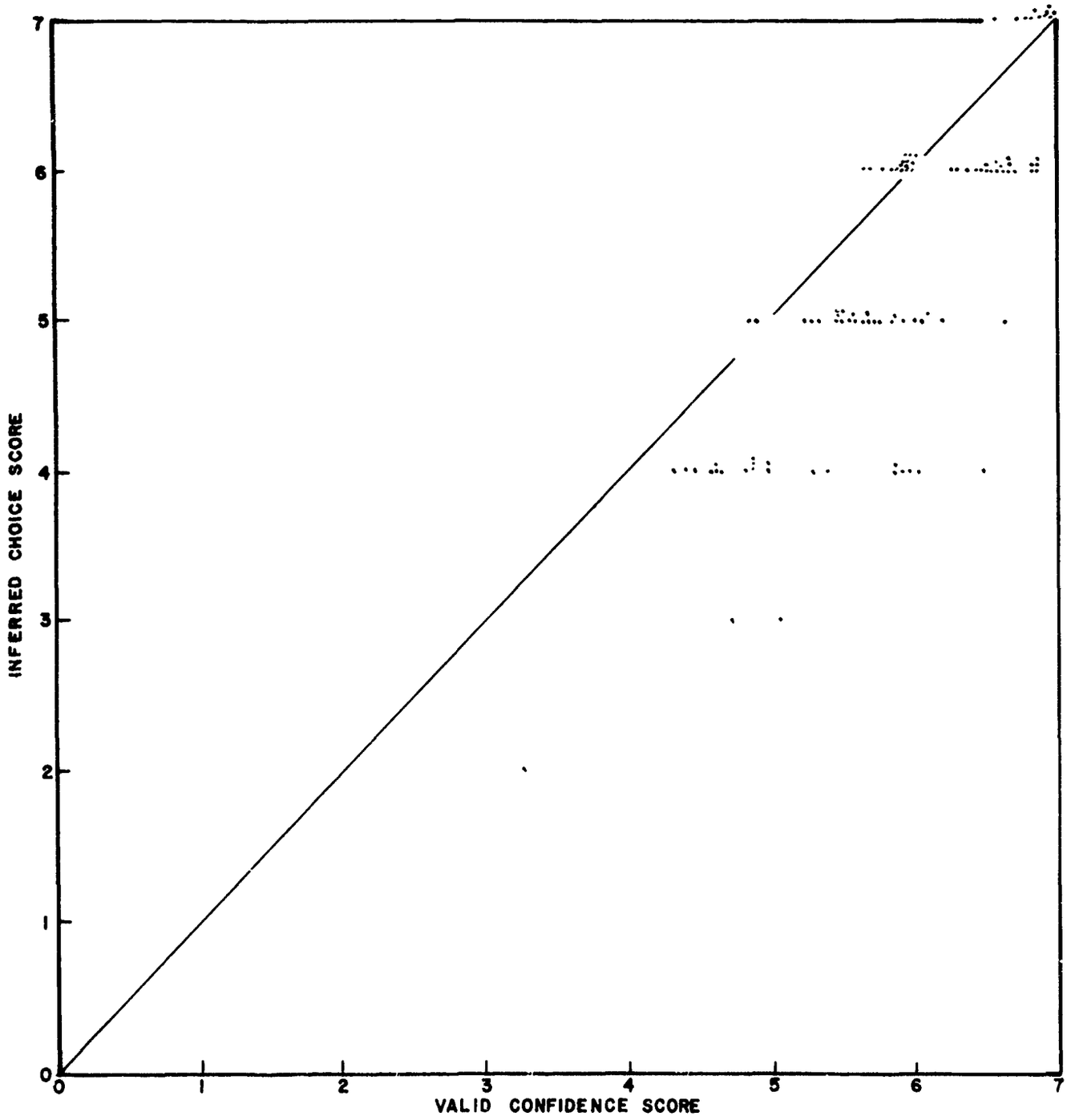

Figure 9. Association between choice and confidence scores for seven item subtest measuring mastery of Teaching Objective No. 2 on L1-2 or L1-2A.

Table 1. Selected Statistics for Each Student

I.	Student Rank on *Valid Confidence* Test		1	2	3	4	5
II.	Total Test Score	A. Inferred Choice Score	47	52	46	46	48
		B. *Valid Confidence* Score Information	52.09	51.95	51.34	50.41	50.41
III.	Number of Items on Which Student is:						
	W - *Well Informed*		34	34	33	38	46
	I - *Informed*		6	17	12	7	2
	P_2 - *Partially Informed* between correct and incorrect answers		10	1	4	3	0
	P_3 - *Partially Informed* between correct and two incorrect answers		2	0	0	0	0
	U - *Uninformed*		1	0	0	0	0
	M - *Misinformed*		4	1	8	8	4
	C - *Completely Misinformed*		1	5	1	2	6
IV.	Misdiagnoses from Choice Data (in Percent)		40	33	41	31	10
V.	Guessing Situations Encountered (in Percent)		22	2	7	5	0

Table 2. Selected Statistics for Each Student

I.		6	7	8	9	10	11	12	13	14	15	16	17	18
II.	A.	49	46	44	44	47	47	47	43	41	42	46	45	43
	B.	50.15	49.87	49.81	49.23	49.14	48.97	48.88	48.57	48.54	48.49	48.41	48.41	48.20
III.														
	W	37	37	35	6	36	30	35	26	18	24	25	40	37
	I	9	9	7	38	10	3	10	14	22	18	16	0	5
	P_2	2	1	4	1	0	9	3	5	2	0	6	10	3
	P_3	1	0	1	1	0	0	0	1	0	0	2	0	0
	U	0	4	3	0	1	0	0	0	1	0	1	0	2
	M	5	1	4	12	7	5	5	10	13	15	6	5	6
	C	4	6	4	0	4	3	5	2	2	1	2	3	5
IV.		29	26	23	90	31	29	31	52	66	57	53	26	28
V.		5	9	14	3	2	16	5	10	5	0	16	17	9

Table 3. Selected Statistics for Each Student

I.		19	20	21	22	23	24	25	26	27	28	29	30	31
II.	A.	44	43	43	45	45	46	41	43	44	45	43	42	45
	B.	48.11	48.08	48.07	48.03	47.93	47.92	47.77	47.77	47.73	47.63	47.62	47.61	47.44
III.														
	W	30	31	19	35	39	41	22	37	39	26	39	24	36
	I	13	5	20	3	1	3	15	0	4	18	4	18	9
	P_2	1	11	6	7	8	4	4	11	3	2	1	2	1
	P_3	0	2	3	1	1	0	1	1	1	0	0	0	0
	U	0	0	0	1	0	0	1	1	0	0	0	1	0
	M	10	6	10	6	1	4	14	2	6	9	12	10	5
	C	4	3	0	5	8	6	1	6	5	3	2	3	7
IV.		44	41	67	31	19	19	60	24	24	50	29	53	26
V.		2	22	16	16	16	7	10	21	7	3	2	5	2

Table 4. Selected Statistics for Each Student

I.		32	33	34	35	36	37	38	39	40	41	42	43	44
II.	A.	46	42	42	44	44	42	41	42	43	43	43	41	44
	B.	47.40	47.32	47.31	47.28	47.26	47.25	47.15	47.15	47.03	47.00	46.96	46.94	46.87
III.														
	W	27	41	35	30	36	14	6	32	35	25	18	37	18
	I	17	0	0	12	8	24	32	4	4	18	24	3	25
	P_2	3	4	13	4	3	5	5	11	4	2	0	4	1
	P_3	0	2	0	1	0	0	1	2	0	0	2	2	0
	U	0	2	2	0	1	3	1	1	6	0	0	2	0
	M	9	4	2	5	2	12	13	4	3	13	12	5	12
	C	2	5	6	6	8	0	0	4	6	0	2	5	2
IV.		50	21	29	38	24	76	90	38	29	57	66	28	66
V.		5	14	26	9	7	14	12	24	17	3	3	14	2

Table 5. Selected Statistics for Each Student

I.		45	46	47	48	49	50	51	52	53	54	55	56	57
II.	A.	44	44	43	41	42	41	44	40	39	41	44	42	42
	B.	46.75	46.64	46.49	46.49	46.39	46.34	46.30	46.24	46.11	46.07	46.03	46.03	46.02
III.														
	W	36	29	28	31	39	9	39	14	20	32	33	22	22
	I	5	12	14	6	0	30	5	23	17	8	9	18	16
	P_2	3	2	1	7	8	1	1	3	3	4	3	2	5
	P_3	1	3	0	2	0	2	0	3	2	0	0	2	2
	U	0	0	0	1	1	3	1	3	0	0	1	2	0
	M	6	9	13	8	3	12	3	10	15	6	3	7	9
	C	7	3	2	3	7	1	9	2	1	8	9	5	4
IV.		26	45	48	41	21	83	17	72	64	31	28	53	55
V.		7	9	2	17	16	10	3	16	9	7	7	10	12

Table 6. Selected Statistics for Each Student

I.		58	59	60	61	62	63	64	65	66	67	68	69	70
II.	A.	41	43	39	42	41	38	42	42	44	40	41	37	42
	B.	46.01	45.91	45.87	45.83	45.81	45.70	45.61	45.42	45.32	45.31	45.19	45.14	44.97
III.														
	W	27	35	31	34	31	26	36	35	36	22	30	22	32
	I	9	6	4	0	9	11	6	7	7	17	11	11	0
	P_2	7	3	7	13	4	3	0	1	0	3	1	6	8
	P_3	0	1	2	1	0	1	0	0	0	0	0	2	3
	U	1	0	2	0	0	0	0	0	1	0	0	1	7
	M	9	7	6	2	9	14	8	9	7	11	8	14	0
	C	5	6	6	8	5	3	8	6	7	5	8	2	8
IV.		45	29	36	28	38	50	24	29	26	53	34	59	31
V.		14	7	19	24	7	7	0	2	2	5	2	16	31

Table 7. Selected Statistics for Each Student

I.		71	72	73	74	75	76	77	78	79	80	81	82	83
II.	A.	42	41	42	36	40	41	40	39	38	36	37	41	35
	B.	44.97	44.81	44.80	44.77	44.73	44.69	44.55	44.16	44.06	43.72	43.72	43.67	43.42
III.														
	W	18	24	35	26	29	33	32	26	24	33	23	35	16
	I	23	12	4	6	8	7	6	12	13	2	12	3	17
	P_2	3	5	5	4	4	0	3	2	2	1	3	7	6
	P_3	0	2	0	6	0	1	0	0	1	0	1	0	0
	U	0	1	0	7	3	1	0	0	1	3	2	0	2
	M	10	14	8	3	10	8	12	15	13	13	14	4	16
	C	4	0	6	6	4	8	5	3	4	6	3	9	1
IV.		62	59	29	45	43	29	36	50	52	33	55	24	71
V.		5	14	9	29	12	3	5	3	7	7	10	12	14

Table 8. Selected Statistics for Each Student

I.		84	85	86	87	88	89	90	91	92	93	94	95	96
II.	A.	37	40	37	39	42	40	40	36	34	31	34	40	35
	B.	43.38	43.28	43.06	42.92	42.52	42.50	42.08	41.91	41.24	41.07	40.55	40.02	39.22
III.														
	W	30	38	23	29	34	37	29	22	16	18	21	30	21
	I	7	0	14	8	6	2	11	10	16	9	10	6	14
	P_2	0	3	1	1	2	1	0	7	4	7	1	3	0
	P_3	0	0	0	2	0	1	0	0	2	0	5	1	0
	U	0	4	0	1	0	0	0	4	1	7	1	0	0
	M	15	0	12	6	9	2	10	7	12	12	8	5	15
	C	6	13	8	11	7	15	8	8	7	5	12	13	8
IV.		38	12	47	31	29	10	36	48	60	60	43	26	50
V.		0	12	2	7	3	3	0	19	12	24	12	7	0

Table 9. Selected Statistics for Each Student

				TOTAL	AVERAGE	PERCENT
I.		97	98			
II.	A.	35	30	4,086	41.69	71.89
	B.	38.68	32.43	4,512.42	46.05	79.39
III.						
	W	22	25	2,851		50
	I	12	0	1,000		18
	P_2	1	9	368		6
	P_3	0	0	77		1
	U	0	0	97		2
	M	14	6	791		14
	C	9	18	500		9
IV.		47	26			41
V.		2	16			10

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
The Shuford-Massengill Corporation One Wallis Court Lexington, Massachusetts 02173	UNCLASSIFIED
	2b. GROUP

3 REPORT TITLE

CONFIDENCE TESTING AT THE OFFICER TRAINING SCHOOL, LACKLAND AIR FORCE BASE: SEPTEMBER 1968

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5 AUTHOR(S) (First name, middle initial, last name)

Emir H. Shuford, Jr. and H. Edward Massengill

6 REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
January 1969	26	3

8a CONTRACT OR GRANT NO F44620-69-C-0068	9a. ORIGINATOR'S REPORT NUMBER(S)
b PROJECT NO 920F-9719	SMC R-14
c 61101D	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d 681313	AFOSR-69-0405-TR

10 DISTRIBUTION STATEMENT

1. This document has been approved for public release and sale; its distribution is unlimited.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research 1400 Wilson Boulevard (SRLB) Arlington, Virginia 22209

13 ABSTRACT

A multiple-choice test on leadership was administered to 98 officer candidates in residence at the Officer Training School, Lackland Air Force Base. Less than one hour was devoted to instructing the officer candidates on how to take a *Valid Confidence* test, and the normal time was then allowed for the students to respond to the 58 test items.

Analysis of the data indicates that taking a *Valid Confidence* test requires no more time than normally allotted to test administration. All the officer candidates understood the instructions and gave confidence responses which yielded more information than choice responses would have. Wide individual differences were observed in the officer candidates' ability to realistically evaluate the quality of information.

Test scores yielded by *Valid Confidence* testing are related to but are not the same as test scores obtained from choice testing. Several lines of reasoning lead to the conclusion that the *Valid Confidence* or information score provides a fairer basis for assessment than does the choice score. Therefore, to the extent that choice testing passes and fails different students than does *Valid Confidence* testing, use of choice testing as the means of assessing and grading students leads to unfair grades and incorrect instructional decisions.

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT		
Valid Confidence testing training multiple-choice tests fairness in assigning grades validity diagnosis						